Term Project. Data-Driven Application

BKMS 1 (Spring 2025)

Prof. Sang-Won Lee

TA. Joobo Shim

VLDB Lab.

2025.05.15(Thu.)

VLDB Lab.

VLDB Lab.

Project Overview

- Data-Driven Application Development

- Our goal is
  - To develop a backend application that utilizes a **DBMS as the core data layer**.
  - To gain hands-on experience in the full lifecycle of database-driven application development
    - ✓ From schema design to data processing and backend integration.
  - To apply domain knowledge from diverse academic backgrounds to design meaningful and impactful data-driven applications.
  - To apply database technologies in data science contexts or creatively integrate AI (such as LLM) with databases to explore innovative ideas.
  - To enhance students' ability to design, implement, and deploy systems that are data-centric and scalable in real-world scenarios.

VLDB Lab.

Project Topics

- RAG-Based Chatbot
  - Implement a chatbot that uses Retrieval-Augmented Generation with a custom PostgreSQL knowledge base.
  - https://medium.com/data-science-collective/rag-in-action-build-your-own-local-pdf-chatbot-as-a-beginner-96c2833869ff

- LLM Integration with MCP (Model-Context-Protocol) + PostgreSQL MCP Server
  - Text2SQL / NL2SQL
  - SQL Debugging
  - SQL Tuning & Optimization
  - Query Result Visualization
  - … LLM + DB

- Any other creative AI + DB Integration project

- You may choose from the list above or propose your own creative idea.

VLDB Lab.

Team

- Team
  - Teams are formed freely, with **3 members** as the standard.
  - Teams may include up to 5 members.
    - ✓ but larger teams are expected to produce higher-quality project
  - A maximum of **2 members from the same lab** is allowed per team.
    - ✓ If more than two members from the same lab wish to work together, they must obtain prior approval from the instructor.

VLDB Lab.

Evaluation Criteria

- Do’s
  - **Creativity & Originality**: How novel and insightful is the idea? Does it tackle a meaningful or underexplored problem?
  - **Domain Insight**: How well does the project apply domain knowledge to define and solve a data-driven problem?
  - **Data-Centric Design**: Is the project truly driven by data (not just UI or functionality)? Is the database a core component?
  - **Technical Appropriateness**: Are the chosen technologies (DB design, backend integration, LLM usage, etc.) applied meaningfully and effectively?
- Don’ts
  - Flashy or unnecessary front-end design
  - Features unrelated to core data logic
  - Overuse of tools without clear purpose
  - Copy-paste projects with no originality

VLDB Lab.

Deliverables

- Source code (per team)
  - Zip the source code of your application. Include all necessary files and directories required to run the application. (Don’t need to include token or access id)
- Report (per team)
  - Free pdf form. The report should include:
    - ✓ Problem Statement & Approach
    - ✓ Project Topic & Description
    - ✓ Database Integration Design
    - ✓ Team Roles
    - ✓ Development Process & Challenges
    - ✓ Lessons Learned & Reflections
- Contribution (individual)
  - The total contribution in the Excel (.xlsx) file must sum to 100 across all team members.
- Late submissions are penalized by 20% of total grade per day.

VLDB Lab.

Schedule

- ~ May 23: Team formation and project topic selection
  - Fill out google form
- ~ June 13: Final presentation
  - 9AM ~ 11 AM at lecture room
  - 5 min presentation (including demo video) + 2 min Q&A
- ~ June 15: Final submission
  - code, report, and contribution